PITCH-JITTER ANALYSIS OF SNORING SOUNDS FOR THE DIAGNOSIS OF SLEEP APNEA

U.R. Abeyratne, C.K.K. Patabandi and K. Puvanendran*
School of Electrical and Electronic Engineering, Nanyang Technological University
Nanyang Avenue, Singapore 639798, E-mail: eabey@ntu.edu.sg
* Sleep Disorders Laboratory, Singapore General Hospital, Outram Park, Singapore

Abstract— **Obstructive Sleep Apnea (OSA) is a disease in which airways involuntarily collapse during sleep, leading to serious consequences. About 10% of snorers suffer from OSA, unknown to them, nevertheless requiring medical attention. The current standard of diagnosis for OSA, polysomnography (PSG), requires that the patients spend one full day in a hospital, wired to a multitude of instruments. PSG is complicated, expensive, and unsuitable for mass screening of the population. OSA is commonly accompanied by snoring. Even though snoring carries vital information on the state of the airways, it has rarely been used in diagnosing OSA. In this paper, we present a mathematical model for snoring, and illustrate its usefulness in diagnosing OSA. We exploit similarities and differences between speech and snoring signals to separate the two, and, provide new features to diagnose OSA at low cost. Via experiments carried out in a hospital sleep-laboratory, we illustrate the importance of using noise reduction techniques to acquire snoring data with sufficient integrity.**
Keywords- **Snoring, Apnea, Pitch Analysis, vocal tract.**

I. INTRODUCTION

An OSA attack is characterized by repeated episodes of upper airway closure during sleep, and, is defined as the total cessation of respiratory airflow that lasts at least 10s. OSA events are typically terminated by a premature arousal from sleep, with the most presenting symptom being loud and interrupted snoring.

Even though OSA appears benign at a first glance, it leads to a large number of untimely deaths. Among the known problems associated with OSA are hypertension, ischemic heart disease and stroke. In addition, OSA is responsible for industrial accidents, driving fatalities and lost production due to daytime sleepiness of operators.

The current standard of diagnosis, PSG, requires that the patients sleeps for a day in a hospital, under video surveillance and wired to a multitude of instruments. In a typical PSG session, signals/parameters such as ECG, EEG, EMG, EOG, nasal/oral airflow, respiratory effort, neck vibrations, body positions, body movements and the blood Oxygen saturation of the patient are carefully monitored. The interpretation of the PSG of a patient too is a complex process, demanding the attention of a trained expert. The limited PSG facilities around the world has resulted in long waiting lists, making it an impossible task to test all the patients in need of such assessment.

There had been a few attempts at using snoring sounds to diagnose OSA [1], [2], [3]. The features commonly used to characterize snoring sounds were the sound intensity or the peak frequency of the snore spectrum. OSA is primarily caused by structural abnormalities in the upper airway during sleep, and the features used did not, unfortunately, directly correspond to OSA. Furthermore, in all of [1], [2], [3], raw snoring data were used without any processing to reduce noise captured together with the data. Even in the controlled setting of the hospital's sleep clinic, the recorded snoring sounds are usually corrupted with background noise, leading to inconsistencies in analysis. One of the serious problems hindering snore analysis is the unavailability of methods to automatically separate genuine snoring sounds from other biological sounds such as somniloquous speech.

This paper addresses those concerns and make the following contributions:

- We show the importance of acquiring snoring signals at a sufficient Signal-to-Noise ratio (SNR), if reasonable results are to be expected. We present methods to enhance the SNR of recordings.
- We propose a mathematical model for snoring, and illustrate its usefulness in separating genuine snoring from other biological sounds such as somniloquous speech.
- We show that the proposed mathematical method allows us to devise novel signatures to diagnose OSA effectively.

II. A model for the snoring

We model the discretized sound $y[n]$ recorded simultaneously with a PSG session as,

$$\begin{aligned} y[n] &= s_s[n] + s_p[n] + b[n], && (1) \\ &= s[n] + b[n] && (2) \end{aligned}$$

where $s_s[n]$ is the clean snoring sound, $s_p[n]$ is the somniloquous speech of the patient, or speech from external sources, and, $b[n]$ is the background electrical and acoustical noise. The quantity $s[n] = s_s[n] + s_p[n]$ now combines clean snoring and somniloquous speech.

The component $s_p[n]$ can be described using the source-vocal tract model in speech synthesis [4], i.e.,

$$s_p[n] = h_p[n] * g_p[n], \tag{3}$$

where $h_p[n]$ represents the transmission characteristics of the vocal tract, and $g_p[n]$ is the source excitation initiating the speech sound. The symbol "*" stands for the linear convolution operator. For voiced sounds such as

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25OCT2001	N/A	-

Title and Subtitle Pitch-Jitter Analysis of Snoring Sounds for the diagnosis of Sleep Apnea	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) School of Electrical and Electronic Engineering, Nanyang Technological University Nanyang Avenue, Singapore 639798	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research Development & Standardization Group (UK) PSC 803 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes Papers from the 23rd Annual International conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 4	

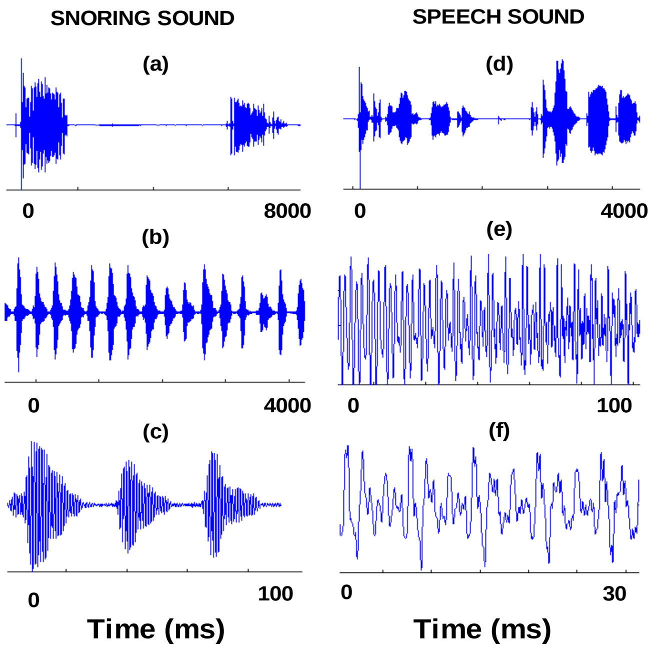

Fig. 1. Speech vs snoring: similarities and differences.

vowels, the excitation can be represented by a sequence of impulses given by,

$$g_p[n] = K \sum_{-\infty}^{\infty} \delta[n - kT_p], \quad (4)$$

where T_p denotes the pitch associated with the particular sound, K is a scaling constant and $\delta[\cdot]$ is the discrete delta function. In the case of unvoiced sounds, $g_p[n]$ is a random noise process. The source-vocal tract model is well developed and is used widely in speech synthesis systems.

In this paper, we draw on the similarities of the speech and snore production mechanism to model the genuine snore component, i.e., $s_s[n]$ in (1). Both snoring and speech sounds share some common "hardware" in the in the process of generation; both are modulated by the vocal tract, or, the acoustical properties of the upper airways.

Some similarities as well as differences between snoring and speech are shown in Fig.1. In Fig. 1 (a), a snoring sample recorded from an OSA patient is illustrated. Fig. 1(b) and 1(c) shows parts of the trace in Fig. 1(a) zoomed into two different scales, for easy visualization of details. Similar plots for a speech signal (corresponding to the sound "ae" in the utterance "one") are shown in Figs. 1(d),(e) and (f). The pseudo-periodic nature of the snoring signal is clearly evident in Fig. 1. This is similar to the case of voiced-sounds in human speech, as seen in Fig. 1(d) (e) and (f).

The "source-excitation" for snoring can be considered to be pseudo-periodic in nature, having, possibly, its origins in the vibrations of the structures of the upper airways. Drawing from the speech model, we describe $s_s[n]$ as,

$$s_s[n] = h_s[n] * g_s[n], \quad (5)$$

where g_s is a source excitation sequence and h_s is termed the Total Airway Response (TAR). The TAR is a slowly time varying function, which captures the time-varying acoustical features of the airways. The quantity $g_s[n]$ is a pseudo-periodic sequence given by,

$$g_s[n] = U[n] \sum_{k=-\infty}^{\infty} \delta[n - kT_s + \epsilon], \quad (6)$$

where ϵ is zero-mean random variable satisfying Probab$(|\epsilon| > T) = 0$, and, $U[n]$ captures the slowly varying magnitudes of the excitation sequence; T_s is the (pseudo) periodicity associated with snoring, which is a measure of the "pitch" of snoring.

In this paper, we investigate the nature of the "pitch" of speech and snoring. Working on clinical data, we illustrate that via Eqs.(3)-(6), we can separate genuine snoring from somniloquous speech, and, more importantly, diagnose OSA consistently.

III. Data acquisition, Annotation & enhancement

A. Data acquisition

The environment of a Sleep Laboratory is highly controlled in order to provide the best ambience for the patient to sleep. However, even in that environment, the component $b[n]$ can drive the SNR of the recording to an unacceptably low value. One of the major reasons is that the component $s_s[n]$ has a dynamic range $> 90dB$. Softer snoring can easily get buried in the background noise $b[n]$. In the work of this paper, SNR enhancement was attempted though careful hardware design, and through software based noise reduction algorithms.

We developed a high fidelity snore acquisition system for the sleep laboratory. Two microphones are used for recordings, one placed about 50cm above the head of the patient, while the other placed near the air-conditioner (see Fig.2). The microphones (40-18000 Hz, Dynamic range 118 dB, Model BG4.1, Shures Brothers Incorporated, Evanston, Illinois) are connected to a signal-conditioning unit (INA103, Burr-Brown Corporation, Tucson, Arizona), output of which is connected to the data acquisition (DAQ) card (NI4552, National Instruments, Austin, Texas) through a shielded coaxial cable of 12m. The sounds are digitized at a rate of 44.1 kSamples/s, with a 16-bit resolution.

In order to achieve a high SNR, we chose low-noise components in the design. Also, we used a shielded and grounded co-axial cable to carry the signal from the examination room to the DAQ in the monitoring station. Shielding proved to be an essential strategy, in countering the electromagnetic interference (EMI) in the environment.

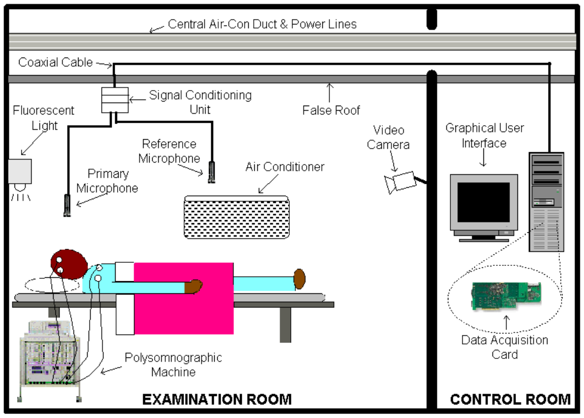

Fig. 2. Snore data acquisition simultaneously with routine PSG assessment.

B. Data Annotation

Based on routine PSG assessments, snoring sounds were annotated as either Benign-Snoring (BS) or OSA-snoring (AS), with the help of clinical specialists. We studied 14 subjects undergoing PSG assessment at the sleep clinic, of which 7 each were diagnosed as belonging to classes AS and BS. About 10 episodes of snoring were randomly selected from each of the subjects for analysis. In patients with OSA, snoring corresponded to the 1st, 2nd and 3rd breath after an episode of OSA attack.

C. Noise reduction

We used the spectral subtraction (SS) method [5] to reduce the background noise component $b[n]$. The SS method was chosen for its simplicity, and the proven ability to deliver high SNR ratios in speech processing applications in noisy environments.

The primary microphone, i.e. the microphone placed just above the patient's head, will capture the signal $y[n]$. We placed another microphone closer to the air conditioning vent to function as a reference microphone, which predominantly captures the soft purring sounds from the air conditioner [6]. The reference microphone output is a reasonable representation of the component $b[n]$, which represents both electrical and acoustical background noise.

Taking the Fourier Transform of (2), we get, $Y(f) = S(f) + B(f)$, where $Y(f)$, $S(f)$ and $B(f)$ respectively denotes the Fourier transforms of $y[n]$, $s[n]$ and $b[n]$.

Then the spectral subtraction can be expressed by:

$$|\hat{S}(f)|^b = |Y(f)|^b - \alpha\overline{|B(f)|^b} \tag{7}$$

where $|\hat{S}(f)|$ denotes the estimate of $|S(f)|$ and $\overline{|B(f)|^b}$ is a measure of the time-averaged noise spectra. Exponent b is set to 1 for magnitude spectral subtraction and 2 for power spectral subtraction. α is the over-subtraction factor and controls the amount of noise subtracted from the noisy signal. To prevent negative estimate, the spectral subtraction magnitude output is further processed as:

$$|\hat{S}(f)| = \overset{½}{\begin{cases} |\hat{S}(f)| & \text{, for } |\hat{S}(f)| > \beta|Y(f)| \\ \beta|Y(f)| & \text{, else} \end{cases}} \tag{8}$$

where the parameter β determines the remaining noise floor. The phase spectrum of the de noised snoring signal is approximated [5] to be the same as the phase spectrum of the noisy snoring signal $\phi_y(f)$. The spectrum of the de noised snoring signal can then be obtained from:

$$\hat{S}(f) = |\hat{S}(f)|e^{j\phi_y(f)} \tag{9}$$

D. The Detection of Pitch

The noise suppressed observation $\tilde{y}[n]$ is used to estimate the pitch associated with snoring, based on a modified version of the well known cepstrum-based pitch detector (Fig. 3). First we computed the envelope $a[n]$ of $\tilde{y}[n]$, and windowed $a[n]$ with $w[n]$ to obtain $a_w[n] = a[n]w[n]$. Then the complex cepstrum $\hat{a}_w[k]$ of $a_w[n]$ was computed according to,

$$\hat{a}_w[k] = F^{-1}\{\log(F(a_w[n]))\}, \tag{10}$$

where F and F^{-1} respectively denote the discrete Fourier Transform and its inverse. The periodicity in $\tilde{y}[n]$ will appear as a peak in $\hat{a}_w[k]$, with the location of the peak corresponding to the period.

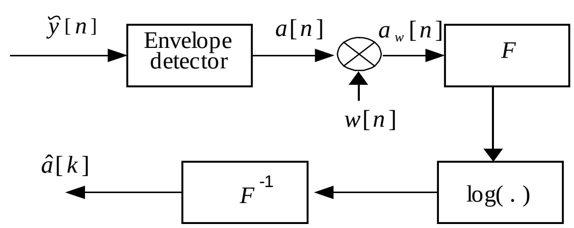

Fig. 3. The modified ceptrum-based pitch detector

IV. Results

In Fig.4(a) we show a 3-second episode of a typical snoring data recorded from a patient. Sound playback of the record confirmed that the background noise profoundly deteriorated the quality of the recording. We used the SS techniques with $b = 1$, $a = 1$ and $b = 0.05$, to obtain the noise suppressed data shown in Fig.4(b). Playback indicated that the SS technique had successfully suppressed the background noise. In general, improvement of SNR in the range of $6 - 8$dB could be achieved, where the SNR is defined as a segmental SNR [4], as used in the context of traditional speech analysis.

Fig.4(c) and 4(d) illustrate the importance of noise reduction in pitch detection. In both frames 4(c) and 4(d), $S(n)$ represents shorter data traces (about 60 ms duration, approx.) extracted from the traces shown in Fig4(a) and 4(b) respectively. In Fig. 4 (c), the raw data

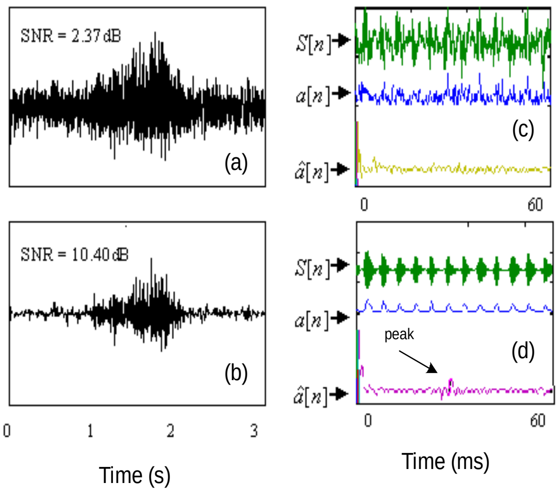

Fig. 4. The importance of background noise suppression.

together with its envelope $a[n]$ and the cepstra $\hat{a}[n]$ are shown. In Fig.4(d), similar figures for noise-suppressed data are shown. Only the de-noised signal shows the periodicity clearly, as also evidenced by the well defined cepstral peak around 30ms. Comparing Figs.4(c) with 4(d), we conclude that noise suppression has an important role in the snore analysis regime. The importance of noise removal takes added significance, if the snore testing is to be done in a home setting, away from the carefully controlled environment of the hospital sleep laboratory.

In pitch analysis of speech, the length of the data window $w[n]$ is taken around 40ms. In snoring analysis, however, this is generally insufficient, because not enough repetitive structures (TAR waves) fall within 40ms for the periodicity to be well detected. Thus, all the results reported in this paper uses a Hamming Window of length 80ms, which proved to be of sufficient length.

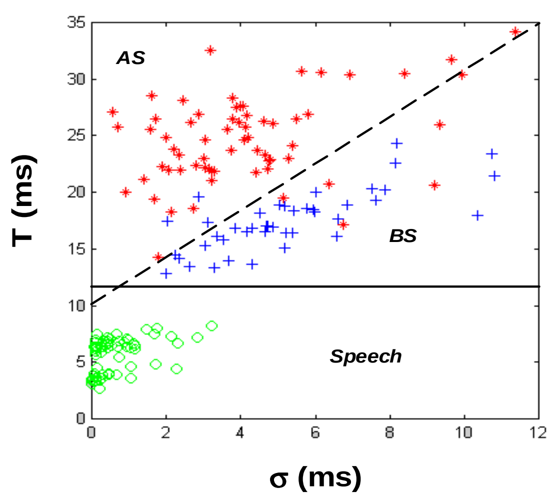

Fig. 5. The pitch-jitter graph for snoring and speech. Symbols '*', '+', and 'o' denote classes AS, BS and Speech respectively.

All the 14 patients in the database were systematically evaluated according to the following procedure:

Step 1: Remove the noise from $y[n]$ using the SS technique to obtain $\tilde{y}[n]$.

Step 2: Define a sliding Hamming Window $w[n]$ of length 80ms, with an overlap factor 50%. Using the cepstral technique, estimate the periods T_i for each segment $i, i = 1, 2, \ldots 30$ of a given snoring/speech episode j as the Hamming window slides over.

Step 3: Calculate the mean $T^{(j)}$ and the standard deviation $\sigma^{(j)}$ for the snoring episode j, based on the 30-estimations T_i, $i = 1, 2, \ldots, 30$. Form the pair $(\sigma^{(j)}, T^{(j)})$ for episode j.

Step 4: Calculate $(\sigma^{(j)}, T^{(j)})$ for all episodes of snoring/speech data belonging to a given class AS, BS or Speech.

In Fig.5, pairs of data $(\sigma^{(j)}, T^{(j)})$ are plotted on a $[\sigma, T]$-plane ('pitch-jitter' graph), where symbols '*', '+', and 'o' respectively denote classes AS, BS and Speech respectively. According to Fig. 5, the snoring signals can be successfully separated into AS and BS classes using the feature $[T, \sigma]$, based on a linear decision boundary $T = 1.85\sigma + 10.0$. This boundary separates the given data into AS class with 92.31% accuracy and BS with 90.7% accuracy. The separation of Speech from the rest of the data was 100% successful.

V. Conclusion

We developed a mathematical model for snoring,in the form of a linear convolution between pseudo-periodic excitation sequence and a quantity (TAR) representing the acoustic-mechanical properties of the upper-airway. We proposed the use of pseudo-periodicity as a signature for OSA. Snoring can easily be discriminated from human speech based on the proposed signature. Furthermore, the pseudo-periodicity itself provides a promising feature to diagnose OSA. Noise reduction schemes are important to obtain good results in snore analysis.

References

[1] D.L. Van Brunt, K.L. Lichstein, S.L. Noe, R.N. Aguillard and K.W. Lester, "Intensity Pattern of Sleep Sounds as a Predictor for Obstructive Sleep Apnea", Sleep, vol. 20, no. 12, pp. 1151-1156, 1997.

[2] J. Rgelio, Perez-Padilla, et al., "Characteristics of the Snoring Noise in Patients with and without Occlusive Sleep Apnea", Am. Rev. of Resp. Dis., vol. 147, pp. 635-644, 1993.

[3] J.A. Fiz, J. Abad, R. Jane, M. Riera, M.A. Mananas, P. Caminal, D. Rodenstein and J. Morera, "Acoustic Analysis of Snoring in Patients with Simple Snoring and Obstructive Sleep Apnea", Eur. Resp. Journal, vol. 9, no. 11, pp. 2365-2370, 1996.

[4] J.R. Deller, J.G. Proakis and J.H.L. Hansen, Discrete-Time Processing of Speech Signals, Upper Saddle River, N.J.:Prentice Hall, 1987.

[5] S.F. Boll, "Suppression of Acoustic Noise in Speech using Spectral Subtraction", IEEE Trans. on ASSP, vol. ASSP-27, no. 2, pp. 113-120, 1979.

[6] T.H.Lee and U.R.Abeyratne, "Instrumentation and Signal Processing for the Detection of Obstructive Sleep Apnea", Proceedings of the European Biomedical.Eng. Conf., Vienna, Austria, 1999.